DOCUMENT RESUME

ED 187 208 HE 012 630

AUTHOR Tierney, Michael L.
TITLE Student College Choice Sets: Toward an Empirical Characterization. ASHE Annual Meeting 1980 Paper.
PUB DATE Mar 80
NOTE 24p.; Paper presented at the Annual Meeting of the Association for the Study of Higher Education (Washington, DC, March 4-5, 1980)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Applicants; *College Bound Students; *College Choice; College Environment; *Decision Making; Higher Education; *Institutional Characteristics; *Socioeconomic Status
IDENTIFIERS *ASHE Annual Meeting 1980; Pennsylvania

ABSTRACT

The nature of student college choice sets was studied. The characteristics of institutions that are assumed to be central to students' application decisions were identified, and students were grouped according to similarity in the types of institutions to which they sent their test scores. Data on 627 students were analyzed by a mode, cluster analytic procedure. The data base was created by merging the American College Testing program and the Educational Testing Service files for students matriculating in the fall of 1975. This data base includes almost 100 percent of the Pennsylvania residents who planned to attend a four-year, degree-granting institution. Two samples of students were selected based upon the socioeconomic characteristics of their residences. One group was drawn from a high socioeconomic area in the Philadelphia metropolitan area, and the second group was selected from a low socioeconomic area in northeastern Pennsylvania. The results indicate that students can be grouped according to their similarity in the types of institutions to which they sent thier test scores. The types of institutions to which students sent their test scores were substantially different according to whether the student was from a low or high socioeconomic status community. (SW)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED187208

# Student College Choice Sets: Toward An Empirical Characterization

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Michael L Tierney

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

By

*Michael L. Tierney*

The Pennsylvania State University
Center for the Study of Higher Education

Presented to

Association for the Study of Higher Education
March 4-5, 1980
Washington, D.C.

HE012630

Abstract

This study investigates the nature of student college choice sets. Data on 627 students in two distinct geographical locations are analyzed by a mode, cluster analytic procedure. The results indicate that students can be grouped according to their similarity in the types of institutions to which they send their test scores. Further, the types of institutions to which students sent their test scores were substantially different according to whether the student was from a low or high SES community.

## Student College Choice Sets: Toward An Empirical Characterization

The Achilles' heel of studies of college choice behavior has been the inability to control for the actual college alternatives considered by students. Lacking such information, investigators have been forced to impute choice sets for potential college students (Kohn et al, 1974; Radner and Miller, 1975). Such imputations are extremely hazardous due to the almost complete lack of information about the institutions to which students apply.

The purpose of this study is to present the results of an initial investigation into the nature of student college choice sets. The study first identifies the characteristics of institutions that are assumed to be central to students' application decisions. The study then groups students according to similarity in the types of institutions to which they sent their test scores. The ability to empirically determine such groups opens a major area of research into the student college choice process.

### Conceptual Framework

The decision to matriculate at a particular institution is the termination of a sequential choice process (Radner and Miller, 1975; Kohn et al, 1974; McPherson, 1978). One possible partition of the student's decision-making problem postulates three stages: (1) the decision as to whether or not to attend a college or university, (2) the decision as to which institutions to apply, and (3) the decision to matriculate at one

institution to which the student has been admitted (assuming the student has been admitted to more than one college). There are, of course, alternative ways of partitioning this complex decision problem (see, for instance, Kohn et al, 1974). Because the manner in which complex decisions are partitioned may affect the overall understanding of the choice process, these alternative partitions also require investigation.

This study will be concerned with the second stage in this college choice process, i.e., the decision to apply to a particular set of institutions. More correctly, this study focusses upon the institutions to which a student sends his/her test score. There is good reason to believe that there is slippage between where a student sends his/her test score and where he/she actually applies. These inconsistencies are also critical for understanding student college choice behavior. A preliminary study of these inconsistencies is the subject of another paper.

A key hypothesis in this investigation is that the nature of institutional alternatives considered by students varies with certain socio-economic (SES) characteristics. Specifically, it is postulated that the SES environment of the student and his/her family establish a set of values and behavioral norms which probably affect the entire college choice process (Trent, 1970). Further, it is expected that the amount of information concerning postsecondary alternatives and the amount of parental influence will vary directly with family and community SES characteristics.

Unfortunately, not much is known about how students generate college alternatives, much less how these decisions are related to SES characteristics. Lewis et al (1974) note that these decisions are largely random, being

dependent primarily upon the haphazard information that students encounter. However, Russick and Olson (1976) found that household income appears to positively affect not only the number of applications sent out by a student but also the amount of parental influence involved in this decision. While suggestive, these studies serve primarily to point to the need for further research into this area.

The study deviates from standard practice by focussing upon the socio-economic characteristics of students within a given geographical (i.e., zip code) area. Implicit in this approach is the assumption that students and families in a given geographical area tend to show the same types of information and to have similar expectations about college, regardless of the parents' actual socio-economic characteristics. A test of this assumption will be undertaken in a more fully developed paper.

For each college to which a student sent an SAT or ACT score, there are a set of characteristics describing that college. It is possible, therefore, to identify certain institutional attributes which may significantly affect the student decision to send a score to a particular institution. Specifically, three institutional attributes are assumed to underlie these decisions.

First and foremost is the geographical proximity of the institution to the student's home. Earlier research has indicated that this proximity may be a critical factor in the decision of students to go on to college (Trent and Medsker, 1968). However, the literature is relatively barren concerning the role of geographical proximity plays in the formulation of student choice sets (Astin et al, 1978).

Two additional institutional attributes were derived from earlier work

by Astin (1962). In this study, Astin found two dimensions tended to differentiate between various institutions; the "wealth" of the institution and its size. Because these two factors tended to differentiate between institutions, it was assumed for the purposes of this study that these differences also differentiate between student choice sets.

Research Method

This study employs data drawn from a recently created sample of 64748 Pennsylvania residents.[1] The data base was created by merging the American College Testing program (ACT) and the Educational Testing Service (ETS) files for students matriculating in the Fall of 1975. This data base includes almost 100 percent of the Pennsylvania residents who planned to attend a four-year, degree granting institution (PDE, 1975).

From this larger data base, two samples of students were selected based upon the socio-economic characteristics of their residences. One group of students (N=163) were drawn from a high socio-economic area in the greater Philadelphia area while a second group (N=464) was selected from a low socio-economic area in northeastern Pennsylvania. This selection process attempted to keep sample sizes relatively manageable for an initial investigation while still maximizing differences in socio-economic environments. Table 1 presents the summary characteristics of these two samples, in comparison with the national norms of college freshmen in the Fall of 1975 (Astin et al, 1975).

These data indicate that the two zip code areas do tend to represent low and high SES regions. Compared to the national norms, substantially more students in the high SES sample are from families with greater than $20,000 annual income and with greater degree aspirations, particularly

Table 1

Sample Summary Characteristics

| Variable | Low SES Sample | High SES Sample | National Norms[1] |
|---|---|---|---|
| Parental Income | | | |
| Less than $3000 | 0.3% | 0.7% | 4.1% |
| $3000 - $6000 | 7.4 | 2.2 | 6.9 |
| $6000 - $15000 | 72.7 | 20.2 | 37.2 |
| $15000 - $20000 | 12.1 | 21.6 | 17.3 |
| Greater than $20000 | 7.4 | 55.3 | 34.4 |
| Race | | | |
| White | 99.3 | 97.2 | 86.4 |
| Non-white | 0.7 | 2.2 | 13.6 |
| Degree Aspirations | | | |
| Associate | 4.3 | 1.2 | 8.0 |
| Bachelor's | 38.5 | 25.4 | 36.0 |
| Masters | 23.9 | 22.0 | 29.4 |
| Doctoral | 16.5 | 28.9 | 9.4 |
| Other Professional | 34.5 | 22.5 | 17.1 |

with respect to the doctorate. Conversely, students from the low SES sample come from families with less than $15,000 annual income. A curious result for this latter sample is the very high percentage of students indicating other professional degrees. Explanations for this result probably include exaggerated expectations, a lack of information as to what these other post-baccalaureate degrees were, and the inclusion by these students of academic programs that are not consistent with the assumed definition of "other professional" degrees.

The three institutional attributes were operationalized as follows. To determine the geographical accessibility of each institutional alternative, the distance between the student's residence and each institutional alternative was calculated. This distance measure was determined using the longitude and latitude coordinates of the student's home residence and the institution's zip code. As such, it is a measure of distance "as the crow flies."

Two variables were developed to operationalize the two primary dimensions along which institutions differ. The "wealth" dimension was measured by the cost to the student of attending a particular institution. More specifically, this cost variable was defined in terms of the tuition and fee charges at each institution as reported on the Higher Education General Information Survey (HEGIS) for 1975-76. In the case where a Pennsylvania resident sends a test score to a public institution in another state, the out-of-state tuition rate is used. The size dimension was measured in terms of a weighted total enrollment estimate as reported on HEGIS for 1972-73.

Summary measures of each student's college choice set are calculated as follows. For a given student, the arithmetic mean for his/her set of institutional alternatives is calculated for each variable. Consequently, if a student applied to three institutions, three scores are calculated to represent the average tuition, size, and distance from home of the actual institutions considered by the student. Similarly, the standard deviation is also calculated in order to assess the amount of institutional variation that occurs in a student's choice set.[2]

Results

Table 2 displays the summary characteristics of student choice sets in these two samples. It is important to note that the measure of dispersion represent the average standard deviation in students' choice sets for a particular group.

First, it is clear that students typically send out three test scores. This result is relatively easily explained by the fact that students are allowed up to three "free" test score reports when taking either the ACT or SAT. This finding is at variance with other studies where a substantial number of students sent out only one application (Corwin and Kent, 1978; Russick and Olson, 1976). One possible explanation for this difference is that the latter studies used actual applications and not test score submission. A second explanation is that the latter studies relied upon student self-reports as distinct from these data which are drawn from ACT and ETS records. In either case, it is apparent that students at this point in the college choice process are involved in a multiple alternative choice problem.

Table 2

Summary Characteristics
of Student Choice Sets

| Variable | Sample: Low SES Area | Sample: High SES Area |
|---|---|---|
| Number of Scores Sent | | |
| None | 16.1% | 24.3% |
| One | 8.6 | 6.2 |
| Two | 15.1 | 11.9 |
| Three | 50.6 | 37.9 |
| Four | 4.5 | 7.3 |
| Five | 2.2 | 4.5 |
| Six | 2.8 | 7.9 |
| Size | | |
| Mean | 8800 | 10214 |
| Average Dispersion | 7926 | 7872 |
| Tuition | | |
| Mean | 1610 | 2091 |
| Average Dispersion | 665 | 601 |
| Distance | | |
| Mean | 139 | 206 |
| Average Dispersion | 120 | 167 |

Second, an examination of three institutional variables indicates that, on the average, students from high SES areas tend to send their scores to slightly more expensive colleges a little further away from home. Such a result would be expected given the folklore surrounding how students apply to college. What is surprising about these averages is how similar low and high SES areas are.

Further, the amount of variation, on the average, in a student's choice set is about the same for both groups, with the minor exception of distances from home. In addition, this variation is large relative to the means for these variables. Such variation would tend to substantiate the folklore that students sent their test scores to a wide variety of colleges, i.e., a "wish" and a "safety" school. The average dispersion in a student's choice set, together with the slight differences in means, would indicate that there is substantial overlap in the choice sets of students in these two samples.

In order to explore this possibility in more detail, a cluster data analytic procedure was applied to the summary characteristics of each student's choice set. The purpose of a cluster analysis procedure is to create groupings of students that are, in some sense, similar to one another. In this particular study, similarity is defined in terms of having similar college choice sets. Initially, the cluster analysis procedure employed is a mode cluster analysis (Wishart, 1969a; 1969b).[3] This procedure was selected because it is appropriate for the identification of "natural" groups within a sample.

Table 3 presents the results for these two samples. Two large clusters were identified for the low SES area. The first cluster is

Table 3

Summary Statistics

for Mode Cluster

Analysis

| Variable | Low SES Area Cluster 1 (N=218) | 2 (N=199) | 3 (N=31) | 4 (N=16) |
|---|---|---|---|---|
| Size | | | | |
| Mean | 3930 | 12974 | 18849 | 3776 |
| Average Dispersion | 1972 | 13975 | 14185 | 2036 |
| Tuition | | | | |
| Mean | 1595 | 1548 | 1478 | 2838 |
| Average Dispersion | 619 | 724 | 689 | 517 |
| Distance | | | | |
| Mean | 89 | 190 | 179 | 98 |
| Average Dispersion | 79 | 165 | 157 | 49 |

Enclosure ratio = .80

Table 3 (Contd)

| Variable | High SES Area Cluster 1 (N=134) | 2 (N=23) | 3 (N=3) |
|---|---|---|---|
| Size | | | |
| Mean | 12038 | 6195 | 2038 |
| Average Dispersion | 9558 | 4036 | 1232 |
| Tuition | | | |
| Mean | 1755 | 2962 | 2636 |
| Average Dispersion | 700 | 356 | 360 |
| Distance | | | |
| Mean | 163 | 348 | 57 |
| Average Dispersion | 145 | 245 | 32 |

Enclosure ratio = .82

characterized by students who tend to send their scores to smaller, moderately priced institutions relatively close to home. These institutions would correspond to the Carnegie Commission (1973) Liberal Arts II institutions which are the dominate type of higher education institution in this area. The second cluster is characterized by larger institutions almost twice as far from the student's home, on the average. These institutions would tend to represent the large public universities in Pennsylvania, most likely dominated by Penn State.

In addition, two small clusters also are identified. The third cluster represents a group which is very similar to the second cluster discussed in the preceding paragraph. It would probably be collapsed with the second group at a subsequent stage in the clustering algorithm. The fourth group includes those students who apply to small institutions relatively close to the student's home. Unlike the first group, however, the cost of attending these institutions is almost twice the cost of the first cluster colleges. Institutions in this cluster would most likely include Bucknell and Lehigh.

The enclosure ratio for this sample is .80, indicating that 80 percent of all the subjects could be identified as belonging to one of the four clusters. Such an enclosure ratio is evidence that clusters identified by this procedure are relatively stable.

Two clusters also dominate student choice sets in the high SES sample. The first cluster in this sample is characterized by large, moderately priced institutions at some distance from the student's home. This cluster appears to be very similar to the second cluster in the low SES sample.

The second cluster represents smaller, much more expensive institutions substantially further away from the student's home. These institutions tend to represent the more elite colleges and universities in the north-eastern United States. The third cluster is so small as to not warrant attention, although a substantive interpretation could be developed. The enclosure ratio for this sample is .82, again indicating relatively stable clusters.

It is apparent that students from both low and high SES areas tend to send their test scores to certain types of institutions, and almost exclusively to institutions having the same cost. Further, the types of institutions considered tend to vary systematically according to the SES area in which the student resides. Thus, students from higher SES areas tend to send their scores to either larger, regional universities or to more elite colleges and universities while students from low SES areas tend to be more location-bound as well as applying to regional institutions. Such findings are extremely suggestive and indicate that this approach to the study of student choice sets may provide substantial insights into college choice behavior.

One possible explanation for these results may be the type of cluster procedure employed. Consequently, a second cluster analysis procedure was applied to the high SES sample. Specifically, Ward's method was utilized (Ward, 1963). This method attempts to identify clusters of students which have minimum within cluster variation. The drawback to this approach is that it assumes that the distribution of data points in this three-dimensional problem is perfectly spherical.

Table 4 presents the results of this analysis. As in the Mode

Table 4

Summary Statistics for Words
Cluster Analysis

| Variable | High SES Area Cluster 1 (N = 109) | 2 (N = 12) | 3 (N = 56) |
|---|---|---|---|
| Size | | | |
| Mean | 12535 | 10986 | 5531 |
| Average Dispersion | 9949 | 8571 | 3641 |
| Tuition | | | |
| Mean | 1709 | 1826 | 2893 |
| Average Dispersion | 726 | 515 | 376 |
| Distance | | | |
| Mean | 114 | 883 | 240 |
| Average Dispersion | 110 | 687 | 166 |

Analysis, Ward's method identified three distinct clusters. Further, while there are some differences in the clusters identified by these two methods, there appears to be substantial similarity in the nature of the clusters identified by these two procedures. In fact, 82.5 percent of the students in this sample are classified into the same groups by these two procedures. A statistical test of the equality of proportions where the samples involve the same students indicates that these two procedures are classifying students into the same clusters (Bishop et al, 1975).

The similarity in results for these two cluster analysis procedures indicates an extremely robust solution. Regardless of differences in the assumptions underlying these two procedures, very similar results are obtained. Thus, not only are the various clusters relatively stable, but they are not affected by the cluster analysis procedure employed.

Discussion

The preceding analyses indicate that is is possible to empirically characterize the choice sets of college students. These characterizations are relatively stable and independent of the cluster analysis procedure employed. The identification of such clusters have substantial implications in at least three broad areas.

A number of research issues are raised about the nature of these clusters. For instance, do these clusters vary according to the nature of the higher education environment? In this study, the low SES area is dominated by private liberal arts colleges. Consequently, students in this area tended to send their scores to those institutions which are geographically accessible. One might hypothesize that at least one cluster in other low SES areas will reflect the nature of the dominant type of

postsecondary institutions in that area. More specifically, one would expect that one cluster for a low SES area in California would be dominated by community colleges.

An additional question is raised concerning the stability of these clusters over time. It will be important to determine whether the nature of these clusters is constant from one high school cohort to another. If a particular region tends to have the same types of clusters across cohorts, it is also necessary to determine whether the same proportion of students in each cohort consider each cluster. Research into these two areas will provide the foundation for the segmentation of the postsecondary education marketplace, and mark a qualitative advancement in enrollment planning.

This paper already provides some evidence that the nature of student college choice sets varies with the community's SES characteristics. As indicated above, it will be important to determine whether or not family SES characteristics are independent of or interact with community SES characteristics in the formulation of a particular choice set cluster. Because the alternatives considered by students varies with community SES characteristics, a number of critical public policy issues are also raised concerning the equality of educational opportunity. These issues will be discussed below.

A second board area suggested by this study concerns the determinants of student college choice sets. More specifically, within a given geographical area, what distinguishes between the students in the various clusters? Preliminary discriminant analyses of the first two clusters in the high SES area indicate that student academic ability, anticipated

college residence, and degree aspirations discriminate among those considering elite and those considering regional institutions. Such results would suggest that the more academically able students are considering more elite institutions while their less academically able but still high SES counterparts are considering regional institutions. Further, this model is able to correctly classify more than 80 percent of the students into the appropriate cluster. One use of such a model is to predict changes in the proportion of students in each cluster given that certain student (or family) attributes are changing within the community. The ability to predict such changes would enhance the enrollment planning capacity of the market segmentation study discussed above.

Finally, the establishment of these clusters raises a number of problems for the provision of equality of educational opportunity. The ability to empirically characterize student college choice sets indicates that students are making critical decisions at this stage in the college choice process. In essence, students are pre-selecting certain types of postsecondary education. This initial "screening" of colleges takes place long before students know whether or not they have been admitted, much less the amount of financial aid they might be offered. Thus, most federal and state policies which are oriented toward providing equality of opportunity (i.e. student financial aid) are not even relevant at this stage in the choice process.

The irrelevancy of student financial aid would be a moot point if students were making fully informed decisions at this stage in the college choice process. However, there are several reasons to believe that this is not

the case. First, a number of students are either not sending out any test scores or only one test score. These students are not even making a choice from among two or more institutions. In fact, one might be led to conclude that these students are constrained by the geographical accessibility of certain institutions. This possibility is especially likely among students from low SES communities.

Second, there apparently are substantial differences among SES areas in terms of the amount of information concerning postsecondary alternatives. One might hypothesize that the reason the more academically able students in the high SES area tend to consider elite institutions rather than regional universities is this information difference. If substantiated by additional research, such results would point to the need for federal and state policies directed at correcting these information differences. Such policies would attempt to provide the "fully informed" consumer which is a necessary condition for the establishment of the equality of educational opportunity.

## References


Astin, A. W., "An Empirical Characterization of Higher Education Institutions." Journal of Educational Psychology, 1962, 53, 224-235.

Astin, A. W., M. R. King, and G. T. Richardson. The American Freshman: National Norms For Fall 1975. Los Angeles: Cooperative Institutional Research Programs, 1975.

Astin, A. W., J. W. Henson, C. E. Christian, G. T. Richardson, and P. E. Hemond, Documentation, SISFAP Study A. Higher Education Research Institute, Los Angeles, California, 1978.

Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland, Discrete Multivariate Analysis. MIT Press, Cambridge, Mass., 1975.

Corwin, T. M. and L. Kent, eds., Tuition and Student Aid: Their Relation to College Enrollment Decisions. American Council on Education, Washington, D. C., 1978.

Kohn, M. G., C. F. Manski, and D. S. Mundel, An Empirical Investigation of Factors Which Influence College-Going Behavior. Santa Monica, California: The Rand Corporation, 1974.

Lewis, G. H., S. Morrison, A. Penz, and B. Wicinas, Nonroutine Decision Making. School of Urban and Public Affairs, Carnegie-Mellon University, Pittsburgh, 1974.

McPherson, M. S., "The Demand for Higher Education." In D. W. Breneman and C. E. Finn (eds,) Public Policy and Private Higher Education. Washington, D. C. The Brookings Institution, 1978.

Pennsylvania Department of Education, Projections: Selected Statistics for Pennsylvania to 1988-89. Harrisburg, Pennsylvania, 1979.

Radner, R. and L. S. Miller, Demand and Supply in U. S. Higher Education. New York: McGraw-Hill Inc., 1975.

Russick, B. and P. M. Olson, "How High School Seniors Choose A College." Choice or Chance: Planning for Independent College Marketing and Retention. Northwest Area Foundation. St. Paul, Minnesota, 1976.

Trent, J. W. The Decision to Go to College: An Accumulative, Multi-Variate Process. Center for the Study of Evaluation, University of California, Los Angeles, 1970.

Trent, J. W., and L. L. Medsker, Beyond High School. San Francisco. Jossey-Bass Inc., 1968.

Ward, J. H. "Hierarchical Grouping to Optimize an Objective Function." Journal of the American Statistical Association, 1963, 58, 236-

Wishort, D. "Numerical Classification Method for Deriving Natural Classes." Nature, 1969, 221, 97-98.

Wishort, D. "Mode Analysis: A Generalization of Nearest Neighbor Which Reduces Chaining Effects." In Numerical Taxonomy. London: Academic Press, 1969.

## Footnotes

1. The creation of this data set was accomplished under an Office of Education grant by the Higher Education Research Institute under the direction of Alexander W. Astin.

2. For each student, the standard deviation for each variable was computed. For a student who applied to only one institution, this standard deviation would be zero. The data reported in Tables 2 and 3 are the average of these standard deviations for the entire sample.

3. Initially, the data analysis plan was to apply the cluster analysis to each student's standard deviations in order to classify students according to the amount of variation in their choice sets. However, only one group was identified by this cluster analysis, indicating that students tend to have the same amount of variation in their choice sets. Consequently, the cluster data analyses reported in this paper are for the mean values of the three institutional variables.